A Kinetic Model of Molecular Titration

William and Mary iGEM 2016

December 1, 2016

Contents

1 Erratum

Until Dec. 1, 2016, a version of this document appeared on this Wiki which contained an incorrect step in the derivation of the kinetic model. This led to a model prediction which, by chance, happened to match the experimental data extremely well. The interpretations and analyses presented in the document were based on this incorrect data.

We realized after the Wiki Freeze deadline that this was the case, and contacted iGEM HQ about the situation. We also informed the judges at the 2016 Jamboree that the modeling document on the Wiki was inaccurate, and that corrected results were used for the Presentation and Poster.

This document now contains the corrected kinetic model, and analyses and interpretations are drawn and conducted appropriately.

2 Motivation

One of the components of our Circuit Control Toolbox is a suite of parts which allow for the use of decoy transcription factor binding arrays that make use of Molecular Titration to shift the sensitivity of a gene or circuit's response to transcription factor concentration. Part of our suite is a series of parts which can be used to construct tetO and lacO binding arrays of arbitrary length using our protocol for UNS-guided Iterative Capped Assembly.

For teams and scientists interested in designing their own decoy binding arrays, it is important to be able to make informed predictions about the part's impact on an existing circuit prior to its actual construction. To facilitate this, we developed a mathematical model which is explicitly parametrized by physiological values for variable aspects of Toolbox parts, such as the number of binding sites on an array or the array's plasmid backbone.

3 Model

3.1 Design and Solution

We developed a simple kinetic ODE model of a basic GFP-expression system, where constitutively expressed tetR inhibits the expression of a pTet-regulated GFP. The model accounts for the activity of aTC, which disables tetR's DNA-binding ability when bound, and the presence of decoy tetO arrays. The kinetic diagram for our model is given by:

$$\emptyset \xrightarrow{\alpha_{tetR}} tetR \xrightarrow{\gamma_{tetR}} \emptyset$$

$$tetR + pTet \underset{k_P^-}{\overset{k_P^+}{\rightleftharpoons}} pTet_{bound} \xrightarrow{\gamma_{tetR}} pTet$$

$$tetR + aTC \underset{k_A^-}{\overset{k_A^+}{\rightleftharpoons}} tetR_{disabled} \xrightarrow{\gamma_{tetR}} aTC$$

$$tetR + tetO \underset{k_O^-}{\overset{k_O^+}{\rightleftharpoons}} tetO_{bound} \xrightarrow{\gamma_{tetr}} tetO$$

$$pTet \xrightarrow{\alpha_{mRNA_{GFP}}} pTet + mRNA_{GFP}$$

$$mRNA_{GFP} \xrightarrow{\gamma_{mRNA_{GFP}}} \emptyset$$

$$mRNA_{GFP} \xrightarrow{\alpha_{GFP}} mRNA_{GFP} + GFP$$

$$GFP \xrightarrow{\gamma_{GFP}} \emptyset$$

which lead to the mass-action ODE system

$$\begin{aligned}
\dot{T} &= \alpha_T - \gamma_(T + TP + TA + TO)T - \\
&\quad k_P^+ T * P + k_P^- TP - k_A^+ T * A + k_A^- TA - k_O^+ T * O + k_O^- TO \\
\dot{P} &= -k_P^+ T * P + k_P^- TP \\
\dot{TP} &= k_P^+ T * P - k_P^- TP \\
\dot{A} &= -k_A^+ T * A + k_A^- TA \\
\dot{TA} &= k_A^+ T * A - k_A^- TA \\
\dot{O} &= -k_O^+ T * O + k_O^- TO \\
\dot{TO} &= k_O^+ T * O - k_O^- TO \\
\dot{m} &= \alpha_m P - \gamma_m m \\
\dot{G} &= \alpha_G m - \gamma_G G
\end{aligned}$$

where for simplicity we have written $T \equiv tetR$, $P \equiv pTet$, $TP \equiv pTet_{bound}$, $A \equiv aTC$, $TA \equiv tetR_{disabled}$, $O \equiv tetO$, $TO \equiv tetO_{bound}$, $m \equiv mRNA_{GFP}$, and $G \equiv GFP$.

Note that we can naturally define conserved quantities in this system which will be invariant over time. These are

$$\begin{aligned}
P_{tot} &= P + TP \\
A_{tot} &= A + TA \\
O_{tot} &= O + TO
\end{aligned}$$

Furthermore, note that one can deduce from the equations that we can intuitively define $T_{tot} = T + TP + TA + TO$. This allows us to define the flux

$$\begin{aligned}
\dot{T_{tot}} &= \dot{T} + \dot{TP} + \dot{TA} + \dot{TO} \\
\dot{T_{tot}} &= \dot{T} - \dot{P} - \dot{A} - \dot{O} \\
\dot{T_{tot}} &= \alpha_T - \gamma_T T_{tot}
\end{aligned}$$

where the second equality is true because for any conservation condition X_{tot}, $\dot{X_{tot}} = 0$.

We are now able to reduce our system of 9 variables down to 6 variables using our conservation conditions:

$$\begin{aligned}
\dot{T_{tot}} &= \alpha_T - \gamma_T T_{tot} \\
\dot{P} &= -k_P^+ T * P + k_P^- (P_{tot} - P) \\
\dot{A} &= -k_A^+ T * A + k_A^- (A_{tot} - A) \\
\dot{O} &= -k_O^+ T * O + k_O^- (O_{tot} - O) \\
\dot{m} &= \alpha_m P - \gamma_m m \\
\dot{G} &= \alpha_G m - \gamma_G G
\end{aligned}$$

This expression can be evaluated at steady-state by setting all of the time derivatives on the

left hand side to 0 and rearranging terms to obtain

$$\begin{aligned}
T_{tot;ss} &= \frac{\alpha_T}{\gamma_T} \equiv J_T \\
P_{ss} &= \frac{K_P P_{tot}}{K_P + T_{ss}} \\
A_{ss} &= \frac{K_A A_{tot}}{K_A + T_{ss}} \\
O_{ss} &= \frac{K_O O_{tot}}{K_O + O_{ss}} \\
m_{ss} &= \frac{\alpha_m}{\gamma_m} P_{ss} \equiv J_m P_{ss} \\
G_{ss} &= \frac{\alpha_G}{\gamma_G} m_{ss} \equiv J_G m_{ss}
\end{aligned}$$

Where $K \equiv k^-/k^+$ is the dissociation constant, and we have defined flux terms J for convenience of notation.

In order to write an expression for G_{ss} in terms of only the rate parameters and conservation conditions of the system, we can see from the above equations that we need to determine an expression for T_{ss} in terms of only rate parameters and conservation conditions. We can do this using the T_{tot} expression: recall that we had defined

$$T_{tot} = T + TP + TA + TO = T + (P_{tot} - P) + (A_{tot} - A) + (O_{tot} - O).$$

This means that

$$T_{tot;ss} = J_T = T_{ss} + (P_{tot} - P_{ss}) + (A_{tot} - A_{ss}) + (O_{tot} - O_{ss}),$$

and we can substitute our expressions for P_{ss}, A_{ss}, and O_{ss} and rearrange terms to eventually obtain

$$0 = T_{ss}^4 + a_1 T_{ss}^3 + a_2 T_{ss}^2 + a_3 T_{ss} + a_4,$$

where

$$\begin{aligned}
a_1 &= -J_T + K_P + K_A + K_O + P_{tot} + A_{tot} + O_{tot} \\
a_2 &= -J_T K_P - J_T K_A - J_T K_O + K_P K_A + K_P K_O + K_A K_O + \\
&\quad P_{tot}(K_A + K_O) + A_{tot}(K_P + K_O) + O_{tot}(K_P + K_A) \\
a_3 &= -J_T K_P K_A - J_T K_P K_O + K_P K_A K_O + P_{tot} K_A K_O + A_{tot} K_P K_O + O_{tot} K_P K_A \\
a_4 &= -J_T K_P K_A K_O
\end{aligned}$$

This provides an expression for which can be solved in a straightforward fashion to compute the value of T_{ss} solely in terms of the rate parameters and conserved conditions of the system. We found that in the parameter values used in this study, T_{ss} always took a single distinct real nonnegative value. Returning to our original reduced model, we see that once we obtain T_{ss} we can follow a cascade of function composition to G_{ss} via

$$G_{ss} = J_G m_{ss} = J_G(J_m P_{ss}) = J_G J_m \frac{K_P P_{tot}}{K_P + T_{ss}}.$$

3.2 Explicit Model Parametrization by Toolbox Parameters

We have now obtained an expression to compute the steady-state concentration of GFP in our system using only the values of the rate parameters and conservation conditions. Note that these conservation condition values correspond explicitly to physiological parameters of toolbox settings in the following way:

The two most easily-tunable parameters for the Titration suite of our Toolbox are the copy number of the plasmid backbone and the number of binding sites on the decoy binding array. We can define two parameters: B to define the mean copy number of the plasmid in a given cell, and C to define the number of copies on a single plasmid. Using the approximation that in an *E. coli* volume, 1 molecule $\approx$ 1 nM concentration [1], we can decompose our conserved quantities as

$$\begin{aligned} P_{tot} &= C_{pTet} * B_{pTet} \\ O_{tot} &= C_{tetO} * B_{tetO} \end{aligned}$$

where we can have distinct values of C and B if the reporter and array are on different plasmids.

This decomposition provides our model with an explicit link between the design specifications of a decoy binding array with its kinetic realization. Details about the reported mean copy numbers for various BioBrick plasmid backbones are given in

`http://parts.igem.org/Help:Plasmid_backbones/Nomenclature`

3.3 Fluorescence to Concentration Conversion

Being a kinetic model, our model processes and evaluates the behavior of the system through concentrations of the relevant proteins. However, we and many users of the Toolbox would likely use fluorescent measurements rather than fluorescent protein concentrations to characterize the behavior of their system. In order to compare our model's results to experimental data, then, we would need a process to convert between fluorescence and fluorescent protein concentration.

Our objective is to obtain an absolute fluorescence unit called MEGFP (Molecules of Equivalent GFP). We modeled our procedure off the workflow developed by the Endy Lab [2] to convert arbitrary fluorescent measurements from a plate reader to MEGFP. We obtained a purified concentration of GFP [3] and diluted it down to various concentration levels and measured these values over different days on the plate reader. The relationship between GFP concentration and GFP fluorescence was linear, as previously described in the literature [4]. We obtained the conversion curve

$$Fluorescence_{MEGFP} = \frac{(Fluorescence_{PlateReader;au} + 683.02)}{100}$$

with an R^2 value of 0.93263. Here $Fluorescence_{MEGFP}$ takes units of concentration (nM).

We also wanted to be able to convert fluorescence values obtained via flow cytometry on our FACS machine to units of MEGFP. To do this, we measured an IPTG induction curve of a standard GFP-expression cassette driven by pLac in the presence of constitutive lacI expression (BBa K2066110) on the FACS machine and converted it to absolute fluorescence units (MEFL) using our standard FACS protocol [see Protocol section of our Wiki]. We then immediately measured these same samples in the plate reader and obtained fluorescence measurements there. We then plotted the plate reader fluorescence values against the FACS fluorescent values at each condition, and found that the the relationship was linear as expected [5]. We obtained the conversion curve

$$Fluorescence_{PlateReader;au} = 0.9054 * Fluorescence_{FACS;MEFL} + 918.10$$

with an R^2 value of 0.9642. This curve can be composed with the above curve to obtain a conversion from MEFL values to MEGFP concentrations.

4 Results

4.1 Model Tuning

In order to validate our model we used it to predict the impact of an 85-repeat tetO array (BBa K2066550) on pSB1C3 backbone on the aTC-induction curve of a pTet-GFP reporter construct in the presence of constitutive tetR expression (BBa K2066053).

We first searched the literature for values for rate constants. We obtained the following values:

Parameter	Meaning	Value	Justification
α_T	Production rate of tetR	? s^{-1}	Fit to Data (see next page)
γ_T	Degradation rate of tetR	0.6207288 s^{-1}	Cell doubling time in M9 minimal media is 67 minutes [6].
K_P	Dissociation constant of tetR with the pTet Promoter	0.1 nM	[7]
K_A	Dissociation constant of aTC with the tetR molecule	0.36 nM	[7]
K_O	Dissociation constant of tetR with the tetO binding site	0.1 nM	[7]
α_m	Transcription rate from the pTet promoter	? nM/s	Fit to Data (see next page)
γ_m	Degradation rate of the GFP transcript	1/120 s^{-1}	Assumption that transcript half-life is 2 minutes [8].
α_G	Translation rate of GFP from transcript	? nM/s	Fit to Data (see next page)
γ_G	Degradation rate of GFP	0.6207288 s^{-1}	Cell doubling time in M9 minimal media is 67 minutes [6].

Table 1: Known Parameter Values for Kinetic Model

Since we were unable to obtain parameter values for the three α terms with which we could be confident, we decided to fit these parameters to an aTC-induction curve of BBa K2066110 without the decoy binding array. We iterated through a parameter range that spanned three orders of magnitude for each of these α terms and minimzed the pointwise squared error between the model-generated steady-state induction curve and the experimentally observed steady-state induction curve. We excluded the last three points in the experimentally-obtained curve from our fitting procedure, as they are likely due to toxicity effects from the high concentration of ATC. The model-generated curves were obtained by solving our model for G_{ss} with $P_{tot} = 1 * 200$ and A_{tot} given by the aTC induction point. We obtained a best-fit parameter set of $\alpha_T = 245.5882$, $\alpha_m = 0.0023$, and $\alpha_G = 125$ (Fig 1).

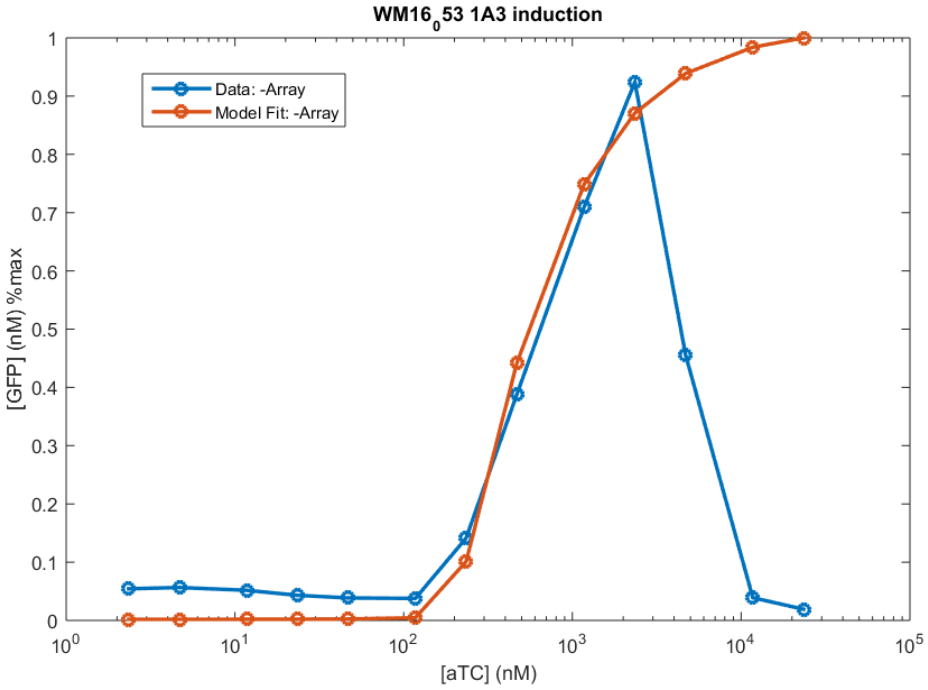

Figure 1: Optimized fit of Kinetic Model to observed induction curve of pTet GFP (BBa K2066053). $\alpha_T = 245.5882$, $\alpha_m = 0.0023$, and $\alpha_G = 125$. $P_{tot} = 1 * 200$ and $O_{tot} = 0$. Curves were fit prior to normalization to the maximal induction condition.

4.2 Assessment of Predictive Power

We then fixed our optimized α parameters and included the 85x tetO array into our model. Because our decoy binding array has 85 repeats, and because pSB1C3 is reported to have 100-300 copies per cell [9], we follow our guidelines from Section 2.2 and set $O_{tot} = 85 * 200$. However, in doing so our model predicted that aTC induction would not occur (Fig 2). This is probably because the amount of decoy binding sites is so great relative to the free tetR concentration that pTet production is already near its maximal expression. Thus the influence of aTC is negligible on the already-induced process. This clearly does not match the measured data, in which we obtained a leftward-shifted induction curve.

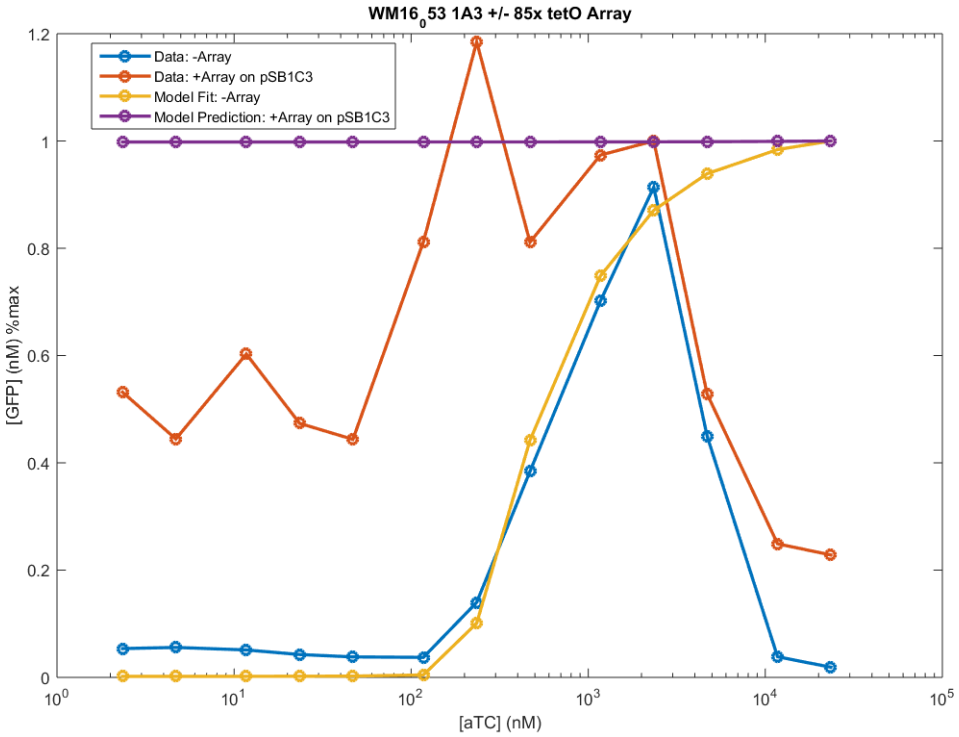

Figure 2: The Kinetic Model does not correctly predict the impact of the 85x tetO array on pSB1C3 on the induction curve of BBa K2066053 on pSB1A3. The +Array model curve was not fit to to the +Array data curve. Curves are normalized to their maximal expression level. α and P_{tot} values are as before. $O_{tot} = 85 * 200$.

In order to investigate the discrepancy between the model-generated response and our experimentally observed response, we simulated a number of different plasmid backbones for the 85x tetO array with the α parameters still fixed as before. We found that a plasmid backbone with mean copy number 5 can replicate the observed normalized induction curve (Fig 3).

5 Discussion

While we have demonstrated that our simple kinetic model is able to generate induction curves which match our experimentally-observed decoy binding array data, we find that the results for our high-copy binding array are only explained by a simulation of a low-copy binding array. A number of factors likely contribute to this discrepancy.

The primary factor is perhaps that the cells from this experiment are likely under greater-than-usual conditions of metabolic stress. This is supported by the fact that high concentrations of aTC seem to be inducing a toxicity effect that depletes gene expression. This may manifest itself in a suppression of the array, through inhibition of its replication or an expulsion of the plasmid. The latter mechanism draws support from our single-cell FACS data of these measurements, in which we observed bimodality in the +Array condition's measurements that we did

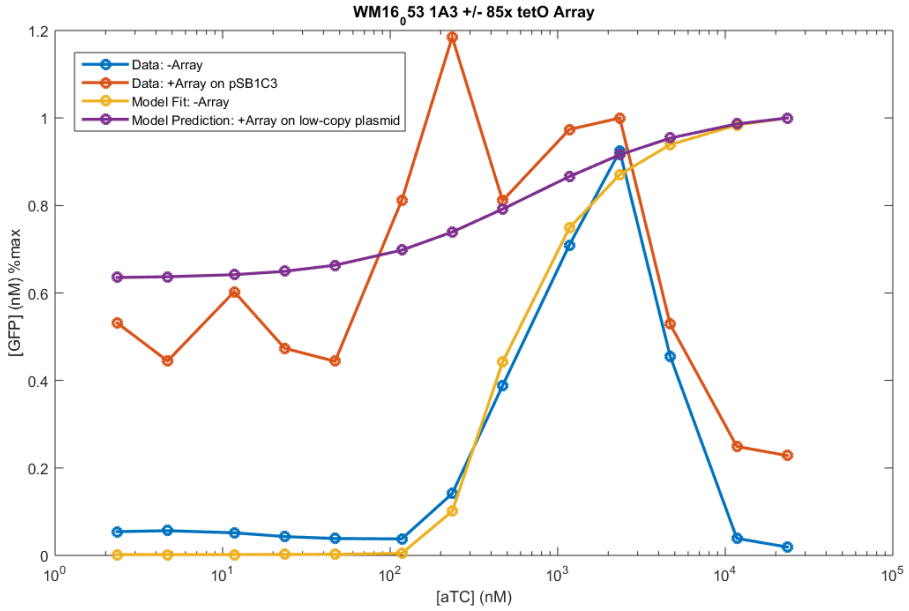

Figure 3: A simulated 85x tetO binding array on 5-copy plasmid backbone replicates the observed induction curve of 85x tetO binding array on pSB1C3. The +Array model curve was not fit to the +Array data curve. Curves are normalized to their maximal expression level. α and P_{tot} values are as before. $O_{tot} = 85 * 5$.

not observe in the -Array condition (Fig 4). This bimodality might be explained by a selection for removal of the array in the presence of insufficient antibiotic selection or selection for a disabling of the array to prevent high protein expression. Hence the observed +Array curve could be the result of a population-level averaging effect between a model-predicted -Array curve an a model-predicted +Array on pSB1C3 curve.

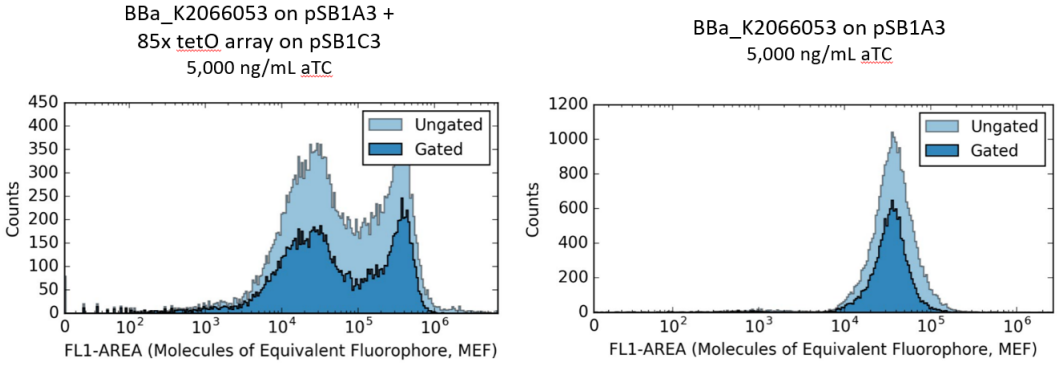

Figure 4: Representative FACS plots of the same induction condition of the BBa K2066053 induction curves with (left) and without (right) the 85x tetO binding array on pSB1C3.

Thus we see that although the decoy binding array can indeed exhibit the expected effects of molecular titration on a circuit component's induction curve, obtaining a precise level of control over the extent of this shift in sensitivity-to-input at the level of binding array design choices such as the number of binding sites or the copy number of the plasmid backbone still remains elusive. Such a level of control would likely require an explicit accounting of the metabolic strain imposed upon the cell by plasmids on high copy numbers. However, this suggests that the precise nature of the impact of low-copy arrays on low-copy reporters or circuits, owing to their lessened strain on the cell, may still be predictable by our simple kinetic model.

References

[1] Milo, R., Phillips, R., & Orme, N. (2015). Cell biology by the numbers. Garland Science.

[2] Kelly, J. R., Rubin, A. J., Davis, J. H., Ajo-Franklin, C. M., Cumbers, J., Czar, M. J., ... & Endy, D. (2009). Measuring the activity of BioBrick promoters using an in vivo reference standard. Journal of biological engineering, 3(1), 1.

[3] `https://www.thermofisher.com/order/catalog/product/88899`

[4] Furtado, A., & Henry, R. (2002). Measurement of green fluorescent protein concentration in single cells by image analysis. Analytical biochemistry, 310(1), 84-92.

[5] Castillo-Hair, S. M., Sexton, J. T., Landry, B. P., Olson, E. J., Igoshin, O. A., & Tabor, J. J. (2016). FlowCal: A user-friendly, open source software tool for automatically converting flow cytometry data from arbitrary to calibrated units. ACS synthetic biology.

[6] `https://www.genome.wisc.edu/resources/k12growth/mg1655growthcurve.htm`

[7] Amit, R., Garcia, H. G., Phillips, R., & Fraser, S. E. (2011). Building enhancers from the ground up: a synthetic biology approach. Cell, 146(1), 105-118.

[8] So LH, Ghosh A, Zong C, Sepúlveda LA, Segev R, Golding I. General properties of transcriptional time series in Escherichia coli. Nat Genet. 2011 Jun43(6):554-60

[9] `http://parts.igem.org/Part:pSB1C3`